

CSE 250B Quiz 7

Tuesday February 25, 2014

Instructions. Do this quiz in partnership with exactly one other student. Write both your names at the top of this page. **Circle one name that we will call out when we return the quiz. Choose a first name or last name that is likely to be unique in the class.**

Discuss the answer to the question with each other, and then write your joint answer below the question. Use the back of the page if necessary. It is fine if you overhear what other students say, because you still need to decide if they are right or wrong. You have seven minutes. The maximum score is three points.

Question. Let α and β be scalar constants. Gibbs sampling for position i for latent Dirichlet allocation uses the result

$$p(z_i = j | \bar{z}', \bar{w}) \propto \frac{q'_{jw_i} + \beta}{\sum_t q'_{jt} + \beta} \frac{n'_{mj} + \alpha}{\sum_k n'_{mk} + \alpha}.$$

Writing $d(j) = \sum_t q'_{jt} + \beta$, this can be rewritten as

$$\begin{aligned} p(z_i = j | \bar{z}', \bar{w}) &\propto (q'_{jw_i} + \beta)(n'_{mj} + \alpha) \frac{1}{d(j)} \\ &= (q'_{jw_i} n'_{mj} + \beta n'_{mj} + q'_{jw_i} \alpha + \beta \alpha) \frac{1}{d(j)} \\ &= \frac{q'_{jw_i} n'_{mj} + \beta n'_{mj}}{d(j)} + \frac{q'_{jw_i} \alpha}{d(j)} + \frac{\beta \alpha}{d(j)}. \end{aligned}$$

As mentioned in class, software can keep track of the $d(j)$ values efficiently. Hence it can also keep track of $\sum_j 1/d(j)$ efficiently. To do Gibbs sampling for position i , we first evaluate the sum over j of each of the three terms above. For the third term, this can be done in constant time. For the first term, evaluating the sum requires $O(A)$ time where A is the number of topics used currently in document m , that is the number of j for which $n'_{mj} \geq 1$. And for the second term, evaluating the sum requires $O(B)$ time where B is the number of topics associated currently with word w_i . We use a random number to select one of these three sums. We then use another random number to choose a topic among those topics that made a nonzero contribution to the selected sum.

Assume that the number K of topics is much bigger than the length of a typical document, and that the constants α and β are small. *Explain why the process above will be faster than standard Gibbs sampling.*

Answer. By the pigeonhole principle, we know that A is much smaller than K . The third sum is selected only when a topic is to be chosen that is not currently used in document m or for word w_i . Because α and β are small, this will happen rarely. Because new topics are selected rarely, we expect B to be small compared to K also. Hence, when one of the first two terms is selected, which is the common case, we need only $O(A + B)$ time to select a new topic.

Additional note. Modeling a large collection of documents, such as Wikipedia, with $K = 1000$ or more topics is reasonable. A typical document and a typical word may involve only ten of these topics, in which case a speedup of close to 100 is achievable. This speedup method was proposed in Section 3.4 of [Yao et al., 2009].

References

[Yao et al., 2009] Yao, L., Mimno, D. M., and McCallum, A. (2009). Efficient methods for topic model inference on streaming document collections. In *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 937–946. ACM.